

BEST AVAILABLE COPY



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03425054.8

CERTIFIED COPY OF
PRIORITY DOCUMENT

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

THIS PAGE BLANK (USPTO)



**Europäisches
Patentamt**

**European
Patent Office**

**Office européen
des brevets**

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03425054.8

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

THIS PAGE BLANK (USPTO



Anmeldung Nr:
Application no.: 03425054.8
Demande no:

Anmeldetag:
Date of filing: 31.01.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

STMicroelectronics S.r.l.
Via C. Olivetti, 2
20041 Agrate Brianza (Milano)
ITALIE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

A reconfigurable signal processing IC with an embedded flash memory device

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G06F15/76

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT SE SI SK TR LI

THIS PAGE BLANK (USPTO)

Title: A Reconfigurable Signal Processing IC with an embedded Flash Memory device.

DESCRIPTION

Field of invention

5 The present invention relates to a dynamically reconfigurable processing unit tightly connected to a Flash EEPROM memory subsystem.

More specifically, the invention relates to reconfigurable signal processing IC with an embedded Flash memory device for non-volatile storage of code, data and bit-streams, the unit being integrated into a single chip together
10 with a microprocessor core.

Prior art

As is well known by those skilled in this technical field, increasing complexity of system design and shorter time-to-market requirements are leading research towards the investigation of hybrid systems including
15 processors enhanced by programmable logic.

In this respect, reference is made to the work by Young-Don Bae et al., "A Single-Chip Programmable Platform Base on A Multithreaded Processor and Configurable Logic Clusters", ISSCC 2002 Digest of Technical Papers, pp 336-337, Feb. 2002.

20 Moreover, a further reference may be considered the article by Zhang et al., having title: "A 1V Heterogeneous Reconfigurable Processor IC for Baseband Wireless Applications", ISSCC 2000 Digest of Technical Papers, pp 68-69,488, Feb. 2000.

At the same time raising costs of mask sets and shorter time-to-market
25 available for new products, are leading to the introduction of systems with a higher degree of programmability and configurability, such as system-on-chip with configurable processors, embedded FPGA and embedded flash memory.

Moreover, the availability of an advanced embedded flash technology,
30 based on NOR architecture, together with innovative IP's, like embedded flash macrocells with special features, is a key factor.

For a better understanding of the present invention reference is also made to the Field Programmable Gate Array (FPGA) technology combining standard processors with embedded FPGA devices.

5 These solutions allows to configure into the FPGA at deployment time exactly the required peripherals, exploiting temporal re-use by dynamically reconfiguring the instruction-set at run time based on the currently executed algorithm.

The existing models for designing FPGA/processor interaction can be grouped in two main categories:

- 10 - the FPGA is a co-processor communicating with the main processor through a system bus or a specific I/O channel;
- the FPGA is described as a function unit of the processor pipeline.

The first group includes the GARP processor, known from the article by T. Callahan, J. Hauser, and J. Wawrzynek having title: "The Garp
15 architecture and C compiler" IEEE Computer, 33(4) : 62-69, April 2000. A similar architecture is provided by the A-EPIC processor that is disclosed in the article by S. Palem and S. Talla having title: "Adaptive explicit parallel instruction computing", Proceedings of the fourth Australasian CComputer Architecture Conference (ACOAC), January 2001.

20 In both cases the FPGA is addressed via dedicated instructions, moving data explicitly to and from the processor. Control hardware is kept to a minimum, since no interlocks are needed to avoid hazards, but a significant overhead in clock cycles is required to implement communication.

25 Only when the number of cycles per execution of the FPGA is relatively high, the communication overhead may be considered negligible.

In the commercial world, FPGA suppliers such as Altera Corporation offer digital architectures based on the US Patent No. 5,968,161 to T.J. Southgate, "FPGA based configurable CPU additionally including second
30 programmable section for implementation of custom hardware support".

Other suppliers (Xilinx, Triscend) offer chips containing a processor embedded on the same silicon IC with embedded FPGA logic. See for

instance the US Patent 6,467,009 to S.P. Winegarden et al., "Configurable Processor System Unit", assigned to Triscend Corporation.

However, those chips are generally loosely coupled by a high speed dedicated bus, performing as two separate execution units rather than being merged in a single architectural entity. In this manner the FPGA does not have direct access to the processor memory subsystem, which is one of the strengths of academic approaches outlined above.

In the second category (FPGA as a function unit) we find architectures commercially known as: "PRISC"; "Chimaera" and "ConCISE".

In all these models, data are read and written directly on the processor register file minimizing overhead due to communication. In most cases, to minimize control logic and hazard handling and to fit in the processor pipeline stages, the FPGA is limited to combinatorial logic only, thus severely limiting the performance boost that can be achieved.

These solutions represent a significant step toward a low-overhead interface between the two entities. Nevertheless, due to the granularity of FPGA operations and its hardware oriented structure, their approach is still very coarse-grained, reducing the possible resource usage parallelism and again including hardware issues not familiar nor friendly to software compilation tools and algorithm developers.

Thus, a relevant drawback in this approach is often the memory data access bottleneck that often forces long stalls on the FPGA device in order to fetch on the shared registers enough data to justify its activation.

The technical problem of the present invention is that of providing a new kind of reconfigurable processing unit tightly connected to a memory architecture having functional and structural features capable to offer significant performance and energy consumption enhancements with respect to a traditional signal processing device.

Summary of invention

The invention overcomes the limitations of similar preceding architectures relying on an embedded device of different nature, and a new approach to processor/memory interface.

According to a first embodiment of the present invention, the reconfigurable processing unit targets image-voice processing and recognition application domains by joining a configurable and extensible processor core and an SRAM-based embedded FPGA.

5 More specifically, the processing unit according to the invention further includes an S-RAM based embedded FPGA unit structured for FPGA reconfigurations having a specific programming interface connected to a port FA of said Flash memory device through a DMA channel.

10 The features and advantages of the processing unit according to this invention will become apparent from the following description of a best mode for carrying out the invention given by way of non-limiting example with reference to the enclosed drawings.

Brief description of the drawings

15 Figure 1 is a block diagram of a processing unit architecture for data processing according to the present invention;

Figure 2 is a block diagram of a Flash memory architecture embedded into the processing unit of Figure 1;

Figure 3 is a schematic view of system memory hierarchy provided by the present invention;

20 Figure 4 is a block diagram of a specific processor extension, for instance added DSP instructions examples;

Figure 5 is a block diagram of a further specific processor extension, for instance an optimized fixed-point calculation of the square root accounts;

25 Figure 6 is a table view showing the overall performance improvements for a face recognition task implemented by the processing unit of the present invention;

Figure 7 is a schematic chip micrograph.

Detailed description

30 With reference to the drawings views, generally shown at 1 is a processing unit realized according to the present invention for digital signal processing based on reconfigurable computing.

The processing unit 1 includes an embedded Flash memory device 4 for non-volatile storage of code, data and bit-streams and a further S-RAM based embedded FPGA unit 3 realized for the configuration purposes of the present invention.

- 5 More specifically, a 8Mb application-specific embedded flash memory device 4 is disclosed. The memory device 4 is integrated into a single chip together with a microprocessor 2 and the FPGA structure 3.

Advantageously, application-specific hardware units are added and dynamically modified by the embedded FPGA 3 reconfiguration. By
10 implementing application-specific vector processing instructions the processing unit 1 shows a peak computing power of 1GOPS.

Efficient read-write-erase access to code, data and FPGA bitstreams is provided by the Flash memory device 4 based on a modular 8Mb, 4-bank Flash memory, as will be more clearly explained hereinafter.

- 15 The processing unit 1 comprises three content-specific I/O ports and delivers an aggregate peak read throughput of 1.2GB/s.

The system architecture 1 is illustrated in Figure 1.

The functional purposes of the embedded FPGA 3 are:

- 20 i) extension of the processor datapath supporting a set of additional special-purpose C-callable microprocessor instructions;
- ii) bus-mapped coprocessors, connected to the system bus through a master/slave interface;
- iii) flexible I/O to connect external units or sensors with application-specific communication protocols.

- 25 Even though such different circuit purposes would require different kinds of programmable logic for best implementation of either arithmetic-dominated or control-dominated logic, a single programmable logic subsystem 3 has been implemented to be shared among different purposes both in space (same configuration) and time (subsequent
30 configurations).

The single, high I/O count, fine-grain e-FPGA 3 operates as a datapath for the microprocessor pipeline and as dedicated control logic for bus

coprocessor and I/O control interface. The FPGA has a specific programming interface 7 connected to a port FP of said Flash memory device 4 through a DMA channel 8.

FPGA reconfiguration is concurrent to software execution.

- 5 A local bus 6 connects a dedicated 32-bit Flash memory port FP to the FPGA programming interface 7.

A DMA channel 8 handles the bitstream transfer while microprocessor fetches instructions and data from different Flash memory ports: 64-bit wide code port (CP) and data port (DP).

- 10 To support streaming applications a 1kB dual-port buffer 9 is used to interface fast decoding hardware and slower software running on the processor 2.

The memory sub-system architecture is shown in Figure 2.

The modular structure of the memory (dotted line) includes:

- 15 - charge pumps 10 (Power Block);
 - testability circuits 11 (DFT);
 - a power management arbiter 12 (PMA); and,
 - a customizable array 13 of N independent 2Mb flash memory modules 16.

- 20 Depending on the storage requirements the number N may be chosen; N=4 in the current implementation.

The modular memory features (N+2) 128-bit target ports and implements a N-bank uniform memory 13.

- 25 As previously mentioned, three content-specific ports are dedicated to code (CP, 64-bit wide), data (DP, 64-bit) and FPGA bit stream configurations (FP, 32-bit). A 128 bit sub-system crossbar 15 connects all the architecture blocks and the eight bit microprocessor 2.

- 30 The main features of such the flash memory device 4 are: charge pump 10 sharing among different flash memory modules 16 through the PMA arbiter 12 in a multi-bank fashion. Moreover, the use of a small eight bit micro processor 2 to easy memory system test and to add complex

functionalities for data management, and the use of an ADC (Analog-to-Digital Converter), required by the application, to increase system self test capability.

5 The third FP port of the Flash device 4 is dedicated to manage embedded-FPGA (e-FPGA) configurations data stored in flash memory modules. The FP port is read-only and provides fast sequential access for bit streams downloading.

The FP has four configuration registers replicating the information stored in CP port that must be used in order to write e-FPGA configurations data.

10 The output data word bus and the address bus are 32 bits wide. The FP port uses a chip select to access in the addressable memory space, and a burst enable to allow burst serial access.

In read operation, an output ready signal is tied low when data are not immediately available, so that it can acts as a wait state signal.

15 The eight-bit microprocessor 2 (uP) performs additional complex functions (defragmentation, compression, virtual erase, etc.) not natively supported by the DP port, and assists for built-in self test of the memory system.

20 The (N+2)x4 128-bit crossbar 15 connects the modular memory with the four initiators (CP, DP, FP and uP) providing that at least three flash memory modules 16 can be read in parallel at full speed.

25 The memory space of the four modules 16 is arranged in three programmable user-defined partitions, each one devoted to a port. The memory system clock can run up to 100MHz, and reading three modules 16 with 128bit data bus and 40ns access time, results in a peak read throughput of 1.2GB/s.

Each 2Mb flash memory module 16 has a 128-bit IO data bus with 40ns access time, resulting in 400Mbyte/s, and a program/erase control unit. Simultaneous memory operations use the power management arbiter 12 (PMA) for optimal scheduling.

30 Available power and user-defined priorities are considered to schedule conflicting resource requests in a single clock cycle.

The memory device 4 allows up to four simultaneous operations, with a limit of one both for write and erase.

Figure 3 depicts the memory hierarchy and parallelism across the processing unit 1. The ports CP and DP are interfaced to the 64-bit, 800MB/s AHB system bus 6.

At a system clock rate of 100MHz each I/O port can independently operate at maximum speed. So, an aggregate peak read rate of 1.2GB/s can be sustained as it is limited by memory access time.

In the current implementation the e-FPGA reconfiguration takes 500 μ s at 100 MHz. 50MB/s average throughput out of the available 400MB/s are currently sustained by the e-FPGA configuration interface 7.

System performance is being evaluated for an image processing application (facial recognition) and a speech recognition application.

More than 20 specific instructions were designed as C/assembly-callable functions, automatically translated to RTL, then synthesized and mapped to the e-FPGA.

Figures 4 and 5 show two examples of specific microprocessor extensions. Figure 4 relates to an eight-issue, eight-bit, L2 calculation accounts for 23 eight-bit arithmetic operations and six 64-bit operations requiring about 10k ASIC equivalent gates.

Figures 5 relates to a datapath for an optimized fixed-point calculation of the square root accounts for twelve 32-bit operations for about 2k ASIC equivalent gates.

The overall performance improvements for the face recognition tasks are shown in the table of Figure 6.

Execution time is compared for 32-bit RISC with basic DSP extensions (MAC, zero-overhead loops, etc) and the same processor enhanced with application-specific instructions.

Measured speed-ups range from 1.8x to 10.6x (on the most-demanding task), with an overall improvement of 8.5x. It must be noticed that switching between algorithm stages requires only one reconfiguration of the e-FPGA. Reconfiguration time is negligible.

The speed-up factors take into account the possible multi-cycle clock penalty due to processor-FPGA synchronization in case of instruction

extensions slower than the processor clock. Energy efficiency figures are reported in Figure 6 too.

As the average power consumption of the system extended with the e-FPGA is slightly higher (10-15%), the energy reduction for executing each
5 of the tasks on its specific HW configuration (power-delay product improvement) results in an overall reduction of 6.7x.

Only one task showed slightly worse total execution energy, though showing benefits on execution speed.

10 Last column of Figure 6 reports the energy-delay improvement of each specific HW configuration compared to the general-purpose counterpart. Energy required for e-FPGA reconfiguration is always negligible.

Measurements show the best energy efficiency in the range of several MOPS/mW at 1.8V supply. It lies between conventional ASIP/DSP and dedicated configurable hardware implementations.

15 The full-processing unit on a single chip is implemented in a 0.18 μ m, 2PL-6ML CMOS embedded Flash technology, chip area is 70mm², technology and device characteristics are summarized in Figure 6 while a chip micrograph is shown in Figure 7.



CLAIMS

1. A dynamically reconfigurable processing unit (1) including an embedded Flash memory device (3) for non-volatile storage of code, data and bit-streams, the unit (1) being integrated into a single chip together with a microprocessor (2) core, further comprising an S-RAM based embedded FPGA unit structured for FPGA reconfigurations having a specific programming interface (7) connected to a port (FA) of said Flash memory device (4) through a DMA channel (8).
2. A dynamically reconfigurable processing unit according to claim 1, wherein said DMA channel (8) handles the bitstream transfer while said microprocessor (2) fetches instructions and data from different Flash memory ports of said Flash memory device (4); a wide code port (CP) and a data port (DP).
3. A dynamically reconfigurable processing unit according to claim 2, wherein said Flash memory device (4) includes a modular array structure (13) comprising N memory blocks (16), and wherein a power block (10), including charge pumps, is shared among different flash memory modules (16) through a PMA arbiter (12) in a multi-bank fashion.
4. A dynamically reconfigurable processing unit according to claim 1, wherein said embedded FPGA unit (3) exploits the following functions:
 - iv) extension of the processor datapath supporting a set of additional special-purpose C-callable microprocessor instructions;
 - v) bus-mapped coprocessors, connected to the system bus through a master/slave interface;
 - vi) flexible I/O to connect external units or sensors with application-specific communication protocols.
5. A dynamically reconfigurable processing unit according to claim 2, wherein said Flash memory device (4) includes at least three different access ports, each for a specific function:
 - said code port (CP) optimized for random access time and the application system;

- said data port (DP) allowing an easy way to access and modify application data; and,
 - said FPGA port (FP) offering a serial access for a fast download of bit streams for an embedded FPGA (e-FPGA) configurations.
- 5 6. A dynamically reconfigurable processing unit according to claim 2, wherein said third port (FP) comprises four configuration registers replicating the information stored in said code port (CP) that must be used in order to write e-FPGA configurations data.
- 10 7. A dynamically reconfigurable processing unit according to claim 5, wherein said third port (FP) uses a chip select to access in the addressable memory space and a burst enable to allow burst serial access.
- 15 8. A dynamically reconfigurable processing unit according to claim 1, wherein said connection between said interface (7) and said port (FA) is provided by a local bus (6).
9. A dynamically reconfigurable processing unit according to claim 5, wherein said Flash memory device (4) includes four modules (16) each arranged in at least three programmable user-defined partitions, each one devoted to a corresponding port.

ABSTRACT

The present invention relates to a dynamically reconfigurable processing unit (1) including an embedded Flash memory device (3) for non-volatile storage of code, data and bit-streams, the unit (1) being integrated into a single chip together with a microprocessor (2) core. Advantageously, the processing unit further comprises an S-RAM based embedded FPGA unit structured for FPGA reconfigurations having a specific programming interface (7) connected to a port (FA) of said Flash memory device (4) through a DMA channel (8).

(Fig. 1)

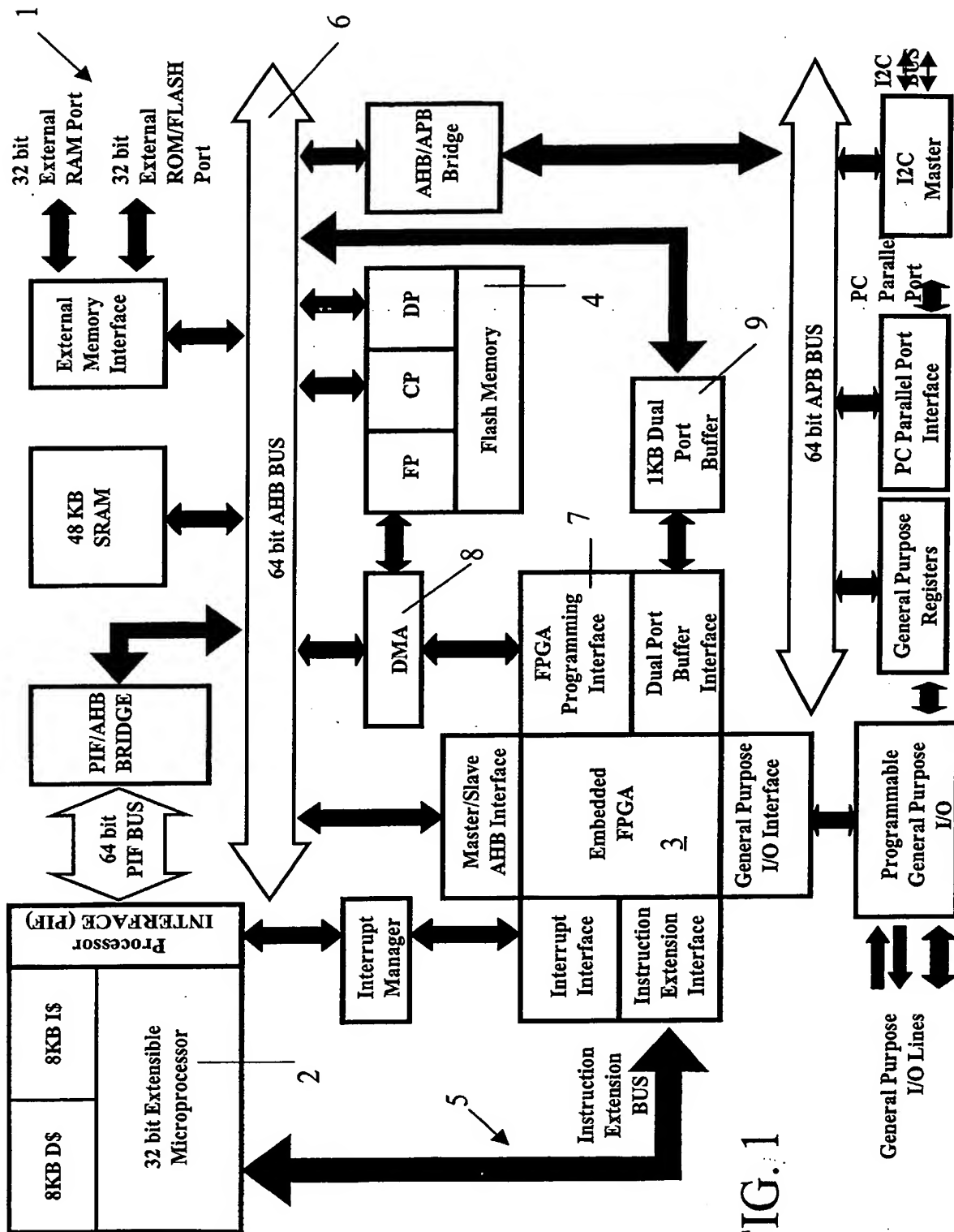


FIG. 1

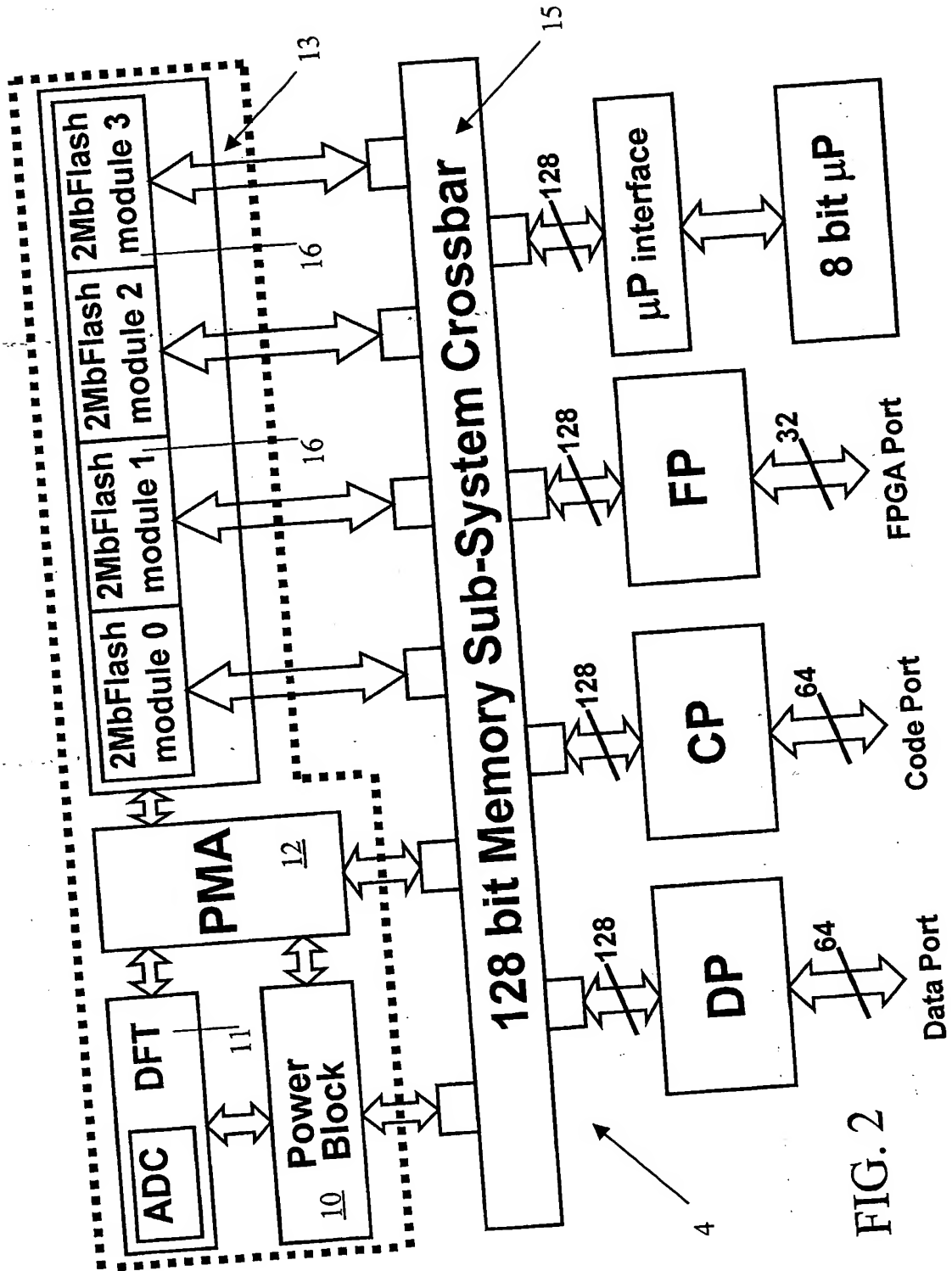


FIG. 2

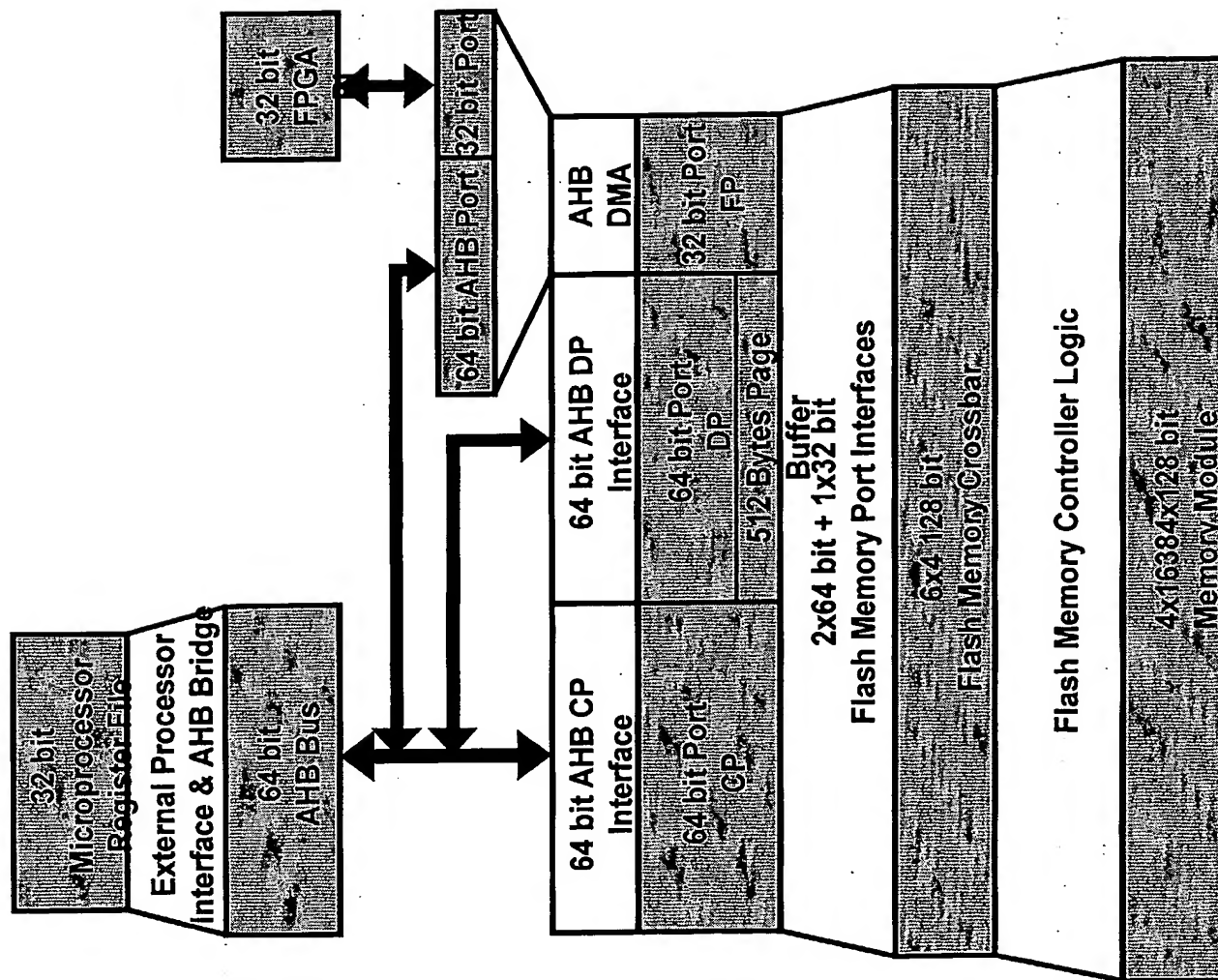
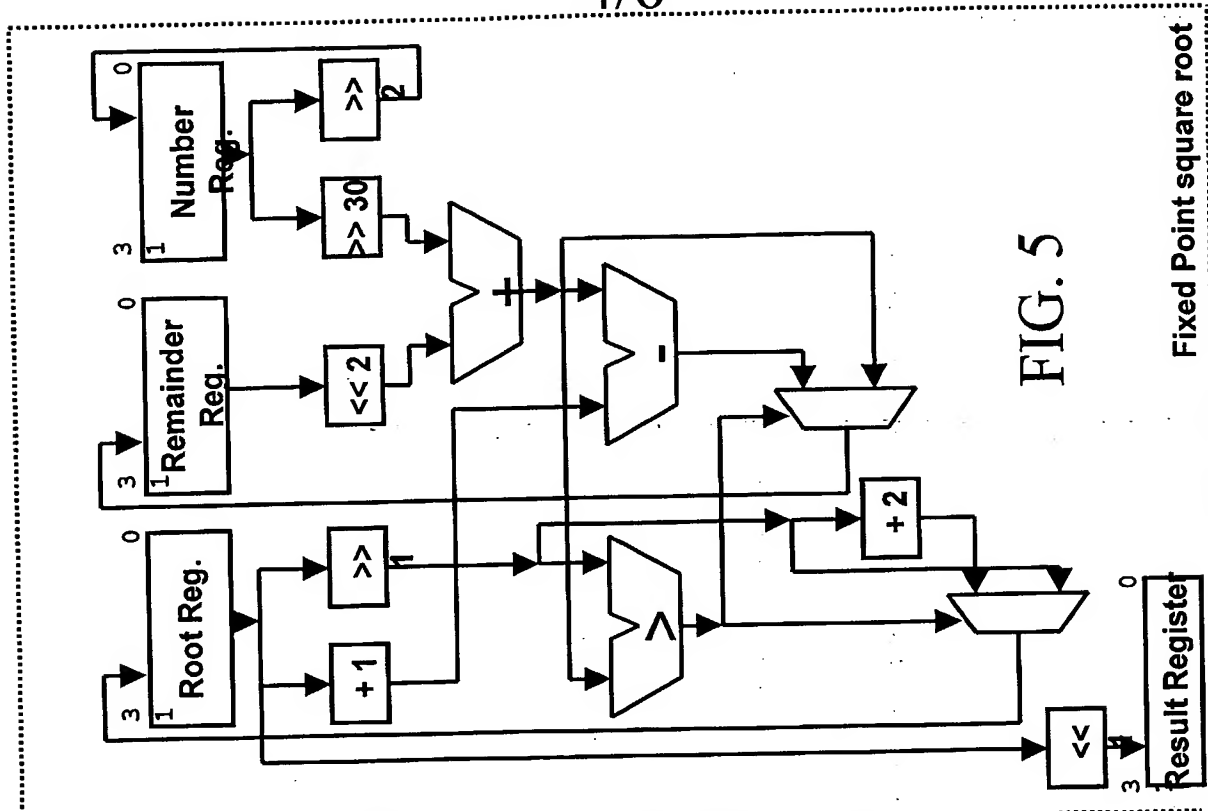
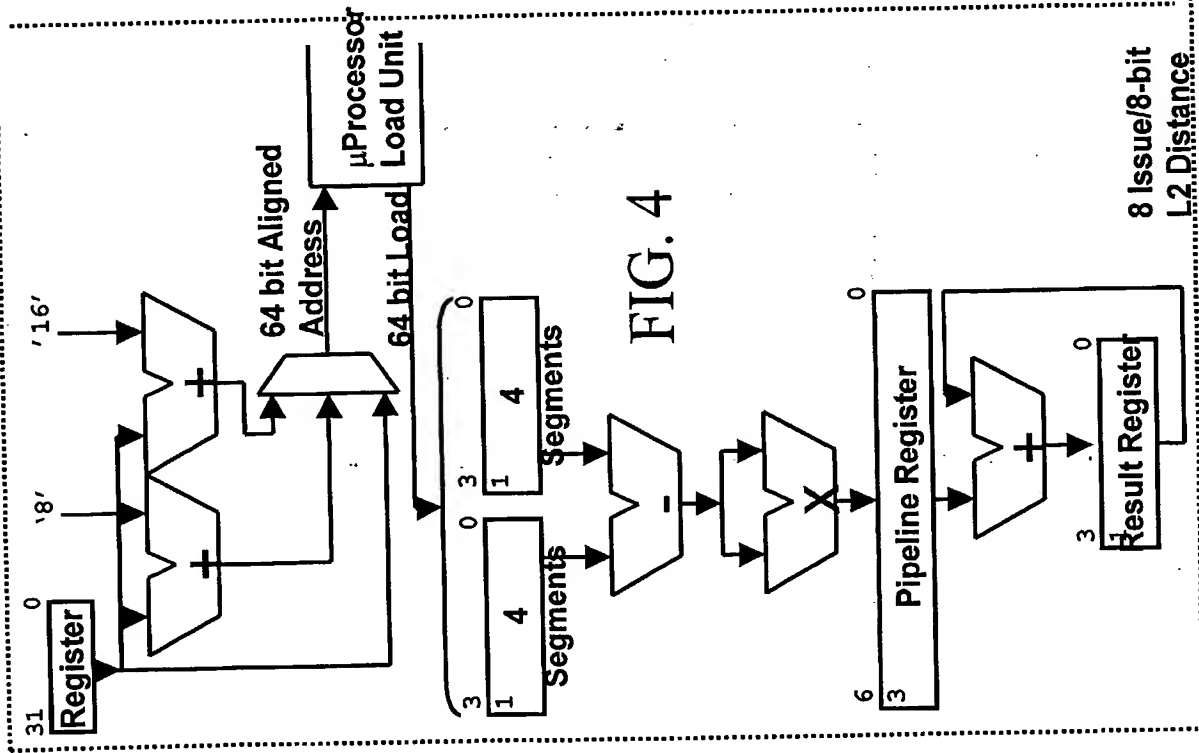


FIG. 3



Algorithm Stage	RISC with basic DSP instruction support	RISC with Microprocessor Extensions	Speed-Up	Energy Gain	Energy Efficiency Gain (Energy x Delay)
Bayer filter	58 msec	24.7 msec	x 2.3	x 1.4	x 3.22
Edge detection	4.5 msec	2.5 msec	x 1.8	x 0.95	x 1.71
Face detection	1.5 sec	382 msec	x 4	x 2.9	x 11.6
Face recognition (Twenty face database)	9.15 sec	860 msec	x 10.6	x 9	x 95.4
Totals	10.7 sec	1.26 sec	x 8.5	x 6.7	

FIG. 6

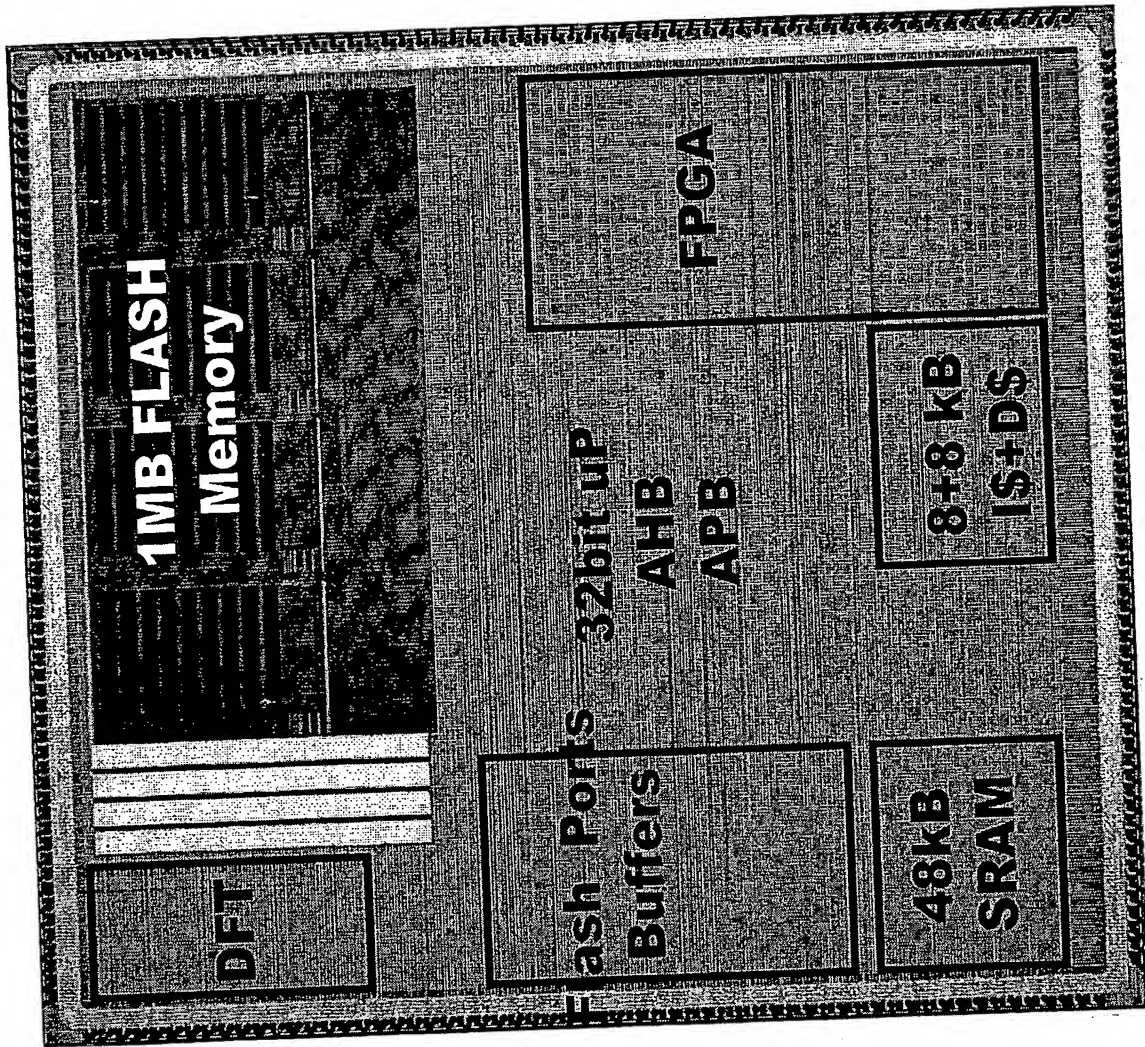


FIG. 7

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKewed/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)